To: Ann Cudd, Dean of Arts & Sciences
From: Mark Crovella, Chair, Department of Computer Science
W. Clem Karl, Chair, Department of Electrical & Computer Engineering
Date: September 8, 2016
Subject: Cross-listing of CAS CS 528 and ENG EC 528

The Department of Electrical and Computer Engineering recently approved a new course, ENG EC 528 Cloud Computing. This course, taught by Professor Orran Krieger, is a formalized version of what had previously been offered as ENG EC 500 / CAS CS 591 Cloud Computing.

Following the lead of ECE, we are requesting that CAS CS 528 Cloud Computing be added to our Department’s course inventory. This course will meet with the newly created ENG EC 528 taught by Professor Krieger, beginning in Spring 2017.

The course syllabus for the most recent iteration of the course is available at http://okrieg.github.io/EC500/index-spring-2016.html#syllabus.

Thank you.
CAS CS 528 / ENG EC 528

Cloud Computing

Description:
Fundamentals of cloud computing covering IaaS platforms, OpenStack, key Big Data platforms, and data center scale systems. Examines influential publications in cloud computing. Culminates in a group project supervised by a mentor from industry or academia.

NOTE: Complete syllabus at http://okrieg.github.io/EC500/index-spring-2016.html#syllabus

Schedule

BU: Tue, Thu 12:00-2:00pm @ PHO211 at BU
NU: Tue, Fri 13:35-3:15pm @ 108 West Willage H (440 Huntington Ave)
Invited talks and Agile Development Lectures: Tue, 2:00-3:00pm @ CAS211 at BU

Announcements

Links

As part of the class, we will be using and referring to:

- Click here for current schedule, this will be changing all the time, so check this frequently.
- Click here for the list of projects.
- Piazza (used for paper reviews and class discussions)

Course Outline

The course will combine group reading and discussion of influential publications in cloud computing, some lectures by instructor and by invited speakers, independent review of talks coupled with classroom discussion, and a large project.

For the most part, the course will be flipped, with most of the material independently studied, and then group discussion of the material.

The project will be done by teams of 3 to 5 students working with a mentor; depending on the project an industry leader and/or developer with a relevant project, or a senior graduate student or a postdoc working on a relevant research project. It will exploit the MOC as well as industry clouds (Amazon AWS, Microsoft AWS, Rackspace, ...)

Course will give students:

- understanding of cloud computing at the IaaS level in general, and open stack in particular
- an understanding of key Big Data platforms
understanding and experience with working as part of an agile team, with experience in running and participating in scrums, planning sessions, ...

- extensive experience with github, agile tools, and various technologies specific to the projects students take on
- an artifact that they will have developed that, if successful, will be part of a broader initiative

**Instructors**

- **Orran Krieger** (okrieg at bu dot edu) Office Hour: Tue 10:00 - 11:00 and Thu 10:00 - 11:00 @ 3 Cummington, room 451,
- Peter Desnoyers (pjd at ccs dot neu dot edu) Office Hours: Tue 10:00 - 11:00 and Thu 10:00 - 11:00 @ 334 West Village H, Northeastern
- **Ata Turk** (ataturk at bu dot edu) Office Hour: Tue 10:00 - 11:00 and Thu 10:00 - 11:00 @ PHO404 or PHO422,
- Michael Daitzman (michael.daitzman at vecna dot com), Office Hour: Thursday 14:00 - 16:00 @ PHO404 or PHO422,
- Jason Hennessey (henn at bu dot edu) Office Hour: Wed 14:00 - 16:00 @ 3 Cummington, room: 553 or 451

**Prerequisites**

This course assumes students have a strong programming background. You will have substantial problems completing the projects if you don't have significant programming experience, and we strongly suggest that you avoid this course, or at least come talk to the instructors if you have any concerns.

**For BU EC students/grads:** Undergrads must have taken EC327 or equivalent and preferably another software course, EC330 or EC440, before taking this course. Graduate students must have taken a rigorous programming class recently, such as EC504 or equivalent (or have major software design experience in industry).

**For BU CS students/grads:** This course assumes students have a strong programming background. Undergrads must have taken CS350 Operating Systems (preferably CS552 as well), CS 460 Database Systems (preferably CS562 as well), and CS 455 Computer Networks (preferably but not must) courses. Graduate students must have taken a rigorous programming class recently, (or have major software design experience).

**For Northeastern students:** This course assumes students have a strong programming background. In particular, MS students should have taken CS 5010, PDP. Industrial programming experience will be helpful, but is not required.

**Workload**

Each week we will be covering on average two research papers and one practitioner paper or other teaching material. You will be expected to read, review and discuss. These papers will likely require that you find and read additional material as necessary to ensure your comprehension. Do not underestimate the amount of work this can be. You will be required to submit a written review of the papers prior to class. You will also be expected to actively participate in the in class discussion. Each student is expected to lead one more of the class
discussions by summarizing the paper and seeding discussions with questions and observations based on the paper.

**Projects**

In addition, to the weekly papers, there will be a major project. This will be teams of 3 to 5 students, starting from the course beginning, with a mentor, from industry, senior graduate student, or one of the instructors.

The project is due by the end of the exam week. The project presentations will be given in the form of a final poster and a demo. There will also be demos scheduled throughout the course to demonstrate regular progress.

The list of projects for this term are:

- Automated Release Notes
- Network traffic collection in the MOC
- Lambda on OpenStack or “Functions as a Service
- Seamless migration between Bare Metal and VMs with OpenStack
- Hybrid cloud services model
- Network Function Containerization
- Container Code Classification
- Learning Analytics
- Cloud Gateway
- Network Programmability as a Cloud Service
- Building a Network Marketplace with XIA
- Flink on Apache Bigtop
- K8PetStore: Deploying a data driven app on Kubernetes with autoscaling
- Cloud Resource Library
- ElasticTest Development
- Cloud and SDN Orchestration
- Adding Trusted Platform Module support to OpenStack
- Cloud Scheduler for Next Generation DataCenters - NOTE this project is mentored at Northeastern
- MOC Monitoring (MOCMon) Platform v2.0
- Magnum Kubernetes Orchestration
- Visualizing MBTA Subway Performance
- Cloud vSwitch Project
- Radiology in the cloud
- Scalable Data Provenance
- Data Provenance and Containers
- OpenDaylight Spectrometer
- Networking as a Cloud Service
- MBTA Alerts and Performance Analysis

There is no final exam.
Grading

- paper discussions, summaries, and reviews (written and verbal) 30%
- project proposal 10%
- bi-weekly project demo and verbal status in aggregate 30%
- final project result and report 30%

Evaluation of project will involve evaluation of demo, poster, code reviews, weekly review of agile progress, and mentor and project partner reviews.

Syllabus

This syllabus is changing as the course progresses, so please check back frequently to see what we will cover next.

The first two weeks we will do course intro, agile methods, and project pitches, as well as some paper reviews. After that, each week we will cover two papers or talks on cloud computing, one discussion topic on agile development or have a speaker about a recent cloud technology, and spend one hour on demos from a subset of the projects. Project teams will be expected to give a demo every second week.

Course calendar: Course syllabus:

Academic Lectures

- You can submit your paper summaries here before first class:
  - Read Cloud computing: state-of-the-art and research challenges (Zhang, Chen & Boutaba)
  - Read Paper Reviewing Guidelines
  - Sign up for Github, Piazza (i.e. reply to invitation)
  - Fill out account and skills survey - http://goo.gl/forms/SUWK238Xze

- Lecture 1
  - lecture - What is Cloud Computing?
  - lecture - paper reviewing guidelines, Project Management Guidelines, project description template, example project description, Project Descriptions
  - Go over Cloud computing: state-of-the-art and research challenges (Zhang, Chen & Boutaba) - PD

- Lecture 2
  - Continued discussion on what is cloud computing
- **The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines** (Only Chapters 1 through 3) - OK

- Lecture 3 - OK
  - Enabling a marketplace of clouds: VMware's vCloud director
  - Toward an Open Cloud Marketplace: Vision and First Steps [Video of talk at OpenStack Summit]

- Lecture 4 - PD
  - Virtualization / Containers [lecture]
  - Disco: running commodity operating systems on scalable multiprocessors
  - Virtualizing I/O Devices on VMware Workstation's Hosted Virtual Machine Monitor
  - Containers - Docker: Introduction to Docker video presentation by Solomon Hykes

- Lecture 5 - OK
  - IX: A Protected Dataplane Operating System for High Throughput and Low Latency
  - SnowFlock: Rapid Virtual Machine Cloning for Cloud Computing

- Lecture 6: MapReduce - AT
  - The Google File System
  - MapReduce: simplified data processing on large clusters

- Lecture 7: Beyond MapReduce - AT
  - Hive - A Warehousing Solution Over a Map-Reduce Framework
  - Spark: Cluster Computing with Working Sets
  - Building a high-level dataflow system on top of Map-Reduce: the Pig experience (Optional reading)

- Lecture 8: The story changes; new networking
  - [Video] Software Defined Networks, video presentation by Scott Shenker
  - Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network
  - B4: Experience with a Globally-Deployed Software Defined WAN

- Lecture 9:
  - Pregel: a system for large-scale graph processing [video (a brief intro to Pregel)]
  - Processing large-scale graph data: A guide to current technology
- **PowerGraph: Distributed Graph-Parallel Computation on Natural Graphs** [video (Joseph E. Gonzales presenting the paper)]

- **lecture 10:**
  - **Flat Datacenter Storage** [video (Jeremy Elson presenting the paper)]
  - **Ceph: a scalable, high-performance distributed file system** [an introductory video]
  - [Also check: CRUSH, data placement algorithm of CEPH and RADOS, underlying object store of CEPH.]

- Northeastern & BU spring break - March 7-11
- **Lecture 11**
  - **Bigtable: A Distributed Storage System for Structured Data**
  - **CAP Twelve Years Later: How the “Rules” Have Changed**

- **Lecture 12:**
  - **Dynamo: Amazon's highly available key-value store**
  - **Cassandra: a decentralized structured storage system**
    - **Introductory blog**
    - **Introduction to Cassandra online course on Datastax**

- **Lecture 13**
  - **Survey of Real-time Processing Systems for Big Data (optional)**
  - **Storm@twitter** [video (Nathan Marz presenting Storm)]
  - **Spark: Discretized Streams: An Efficient and Fault-Tolerant Model for Stream Processing on Large Clusters** [video (Tathagata Das presenting the paper)]

- **Lecture 14:**
  - **Mesos: A Platform for Fine-Grained Resource Sharing in the Data Center** [video (Matei Zaharia presenting the paper)]
  - **Large-scale cluster management at Google with Borg**

- **Lecture 15:**
  - **Kafka: A Distributed Messaging System for Log Processing**
  - **Spanner: Google's Globally-Distributed Database**

- **Lecture 16:** Putting it all together into context
- **Additional reading:**
  - **ZooKeeper: Wait-free coordination for Internet-scale systems** [video (Benjamin Reed presenting the paper)]
The Turtles Project: Design and Implementation of Nested Virtualization [video (Muli Ben-Yahuda presenting the paper)]

Agile Development Lectures

- Agile 1: [BU CAS-211]
  - Before class watch the videos for A history of development and Intro to Scrum - https://youtu.be/RJaF4owQDgg (YouTube lets you play this faster, you may need to use html5 playback, details here: http://lifehacker.com/5993156/watch-lectures-in-half-the-time-with-youtubes-html5-player)
  - Read http://billmoyers.com/content/how-to-tell-your-story-of-self/
    - Jot some notes about yourself for a narrative exchange you will do with your team members
    - if you find yourself getting stuck look at the longer version
      here:http://www.wholecommunities.org/pdf/Public%20Story%20Worksheet07Ganz.pdf for a list of questions to help you think about your narrative
  - before class: checkout: The Agile Manifesto, Context for the Agile Manifesto, Core Scrum, Task Boards, User Stories,
  - Areas we will cover: Story creation using roles, Definition of ready, Sizing stories/Levels of effort, Hours vs Velocity, The Whole Team Approach, Risk management
    - Populating Backlogs, Backlog Grooming, Incremental Development
    - Burndown Charts, Agile Scheduling
  - Using Trello - the tool we will use for your Backlog - Guest lecture Michael Daitzman

- Agile 2: [BU CAS-211]
  - Before the class check out - Horizontal and Vertical User Stories
  - Questions you may have, Ways to speed up getting your team to “Performing”
  - Laura - MOC walkthrough

- Agile 3: [BU CAS-211]

Invited talks:

- Jan 19th 2-3PM Invited talk [BU CAS-211]

  Title: The Massachusetts Green High Performance Computing Center
Abstract: The MGHPCC is a high performance computing datacenter built to support the research computing needs of five Boston area universities: MIT, BU, Northeastern, Harvard and UMass. The center is a public/private venture and was constructed with funds from the five universities, the Commonwealth of Massachusetts, Cisco and EMC. This talk present the major infrastructure components of the facility and explain some of the challenges operating a modern, energy-efficient, high-density computer facility.

Presenter: Jim is the Director of IT Services at the Massachusetts Green High Performance Computing Center. The center is a 10MW computer facility located in Holyoke Massachusetts, an ideal location for computing in Massachusetts with ample land, inexpensive power and access to several high-capacity optical networks. Jim has led the effort to make sure the facility meets the computing and security requirements of the member institutions as the project has progressed through the design, construction, startup and operations phases.

Prior to MGHPCC Jim worked as a consultant to British Telecommunications, CTO at Weather Services International and CTO at MetraTech Corporation. Jim has over 20 years of industry experience as a technology executive, project lead and, operations manager.

- Mar 22nd 2-3PM Invited talk [BU CAS-211]

Title: Intro to OpenStack Software Platform and Community

Abstract: "The talk introduces the audience to OpenStack, the fastest growing open source project to date, and the preeminent open source cloud operating system. The talk will provide an overview of OpenStack, the various components involved in running a cloud, and how the various components communicate amongst themselves. We describe architectural considerations in designing OpenStack that allow it to be extensible and allow for rapidly scalable development processes. As the OpenStack community has exploded in size, certain processes and procedures were put in place in order to allow for more effective governance and management of the myriad projects that comprise the OpenStack ecosystem. Tooling was developed around blueprint and design development, source code management, code reviews, and team communication, all with the goal of providing more consistent and transparent management of project direction. However, with the introduction of these many processes and tools came a burden to new contributors as well as a need for mentorship programs to nurture and guide individuals and organizations into being good OpenStack community citizens. This talk also provides a guide to highlight these processes and tools to help new folks get introduced to the community. We will be covering the OpenStack community and governance tools we use in the community Review and code best practices, Tips and techniques for best interacting with community members etc."

Bio:
Amrith Kumar is an active technical contributor to the OpenStack project, and a member of the Trove core review team. He is also a member of the OpenStack Foundation Job Analysis Task Force and an author of the book on OpenStack Trove. He is the author of
the book on OpenStack Trove (published by Apress, http://www.apress.com/9781484212226). He is the founder and CTO of Tesora, and brings more than two decades of experience delivering industry-leading products for companies specializing in enterprise storage applications, fault tolerant high performance systems and massively parallel databases to Tesora, which he co-founded. Earlier, he served as vice president of technology and product management at Dataupia, maker of the Satori Data Warehousing platform, and Sepaton’s director and general manager where he was responsible for the development of the core virtual tape library product. As a director of product development at Netezza, he managed end-to-end product delivery for all customers and prospects. Amrith studied mathematics at the University of Madras (India) and management at the Indian Institute of Management.

Davanum Srinivas is a long time contributor to OpenStack, currently serves as the Project Technical Lead for Oslo. He is active in several other OpenStack projects like Nova, KeyStone, Magnum and helps with the OpenStack Release team as well. At Mirantis, he leads a team working on improving KeyStone project and infrastructure components like RabbitMQ, MySQL Galera through Oslo projects like Oslo.Messaging and Oslo.DB.

- **April 4, 2-3PM Invited talk [BU CAS-211]**

Jonathan Appavoo described project **Kittyhawk**

**Abstract:** Kittyhawk represents our vision for a Web-scale computational resource that can accommodate a significant fraction of the world’s computation needs and enable various parties to compete and cooperate in the provisioning of services on a consolidated platform. In this paper, we explain both the vision and the system architecture that supports it. We demonstrate these ideas by way of a prototype implementation that uses the IBM Blue Genet/P platform. In the Kittyhawk prototype, we define a set of basic services that enable the allocation and interconnection of computing resources. By using examples, we show how higher layers of services can be built by using our basic services and standard open-source software.

- **April 19th, 2-3PM Invited talk [BU CAS-211]**

**Title: Current state of private cloud networking**

**Abstract:** Summary of cloud industry including public cloud space of AWS, Google Cloud, Microsoft Azure, Rackspace; private cloud space: VMware vSphere, OpenStack. Who are the Users of the cloud? Private cloud networking: VMware NSX, Cisco ACI, Openstack networking. Competitive landscape of private cloud networking: Juniper Contrail, Midokura, Plumgrid, Dragonflow, Nuage. Technical architecture, pros and cons of each system. Things in the future: bridging between multiple clouds, compatibility with carriers, technical features in the future.
Bio: Richard Wong is a Technical Sales Engineer at Midokura, the open-source virtual networking company a leading contributor in the Neutron and Kuryr networking projects. He has deep systems engineering experience, with a focus on systems architecture, network design, and datacenter planning. Richard has worked in a wide variety of platform environments, including VMware, OpenStack, RHEL, Solaris, Windows, AIX, Ubuntu, and has a wide background in SAN storage as well as Cisco networking. He is native of a New York City and have worked at firms such as Cantor Fitzgerald, AIG, DTCC, Dell, Egenera.

- April 26th, 2-3PM Invited talk [BU CAS-211]

Title: Unpacking the box: The evolution of both the cloud data center network and Big Switch Networks, too

Abstract: The requirements placed on the network within the data center have changed rapidly over the last ten years, with solutions evolving to L2/L3 SDN fabrics. A quick survey of the factors that drove the change, the technology that enables hyperscale architecture solutions, the advantages SDN fabrics have and Big Switch Networks’ journey through it all.

Bio: Paul Zablotski is a Systems Engineer at Big Switch Networks. He has more than 20 years’ experience in networking and compute, including previous roles at Juniper Networks, Enterasys Networks, Dunkin' Brands and Fore Systems/Marconi Communications. He's attained many IT industry certifications, though none is more prized than JOAT-MOS.

Project deadlines:

- Projects will be assigned by Thursday Jan 21, for both Northeastern and BU
  - Send an email to your mentor, schedule a meeting (physical or virtual) within next week
  - Hold a team meeting before Tuesday's class

Programming Assignments:

- Assignment 1 (Due 2/4):
  - Programming tutorial assignment on Docker
  - Follow the getting started tutorial. Choose the tutorial for your OS.
  - Write a one paragraph summary of your experience with the tutorial, submit your summary. We will grade these summaries.

- Assignment 2 (Due 2/16):
  - Assignment on Agile Practices
  - Watch Michael's video on First Class 2016 with Audio3
Write a one paragraph summary for video, submit your summary. We will grade these summaries.

Assignment 3 (Due 2/18):

- Programming tutorial assignment on Spark
- Follow the hands on with Spark tutorial.
- Write a one paragraph summary of your experience with the tutorial, submit your summary. We will grade these summaries.

Assignment 4 (Due 3/17):

- Programming tutorial assignment on Hadoop, Pig, and Hive: Follow the instructions:
  - Install Virtualbox (https://www.virtualbox.org) if you haven’t done so already
  - Download and install the Hortonworks Sandbox image for Virtualbox from http://hortonworks.com/products/hortonworks-sandbox/#install
  - Go through the following tutorials:
  - Write a one paragraph summary of your experience with each tutorial, submit your 4 paragraph summary. We will grade these summaries.

Assignment 5 (Due 3/31):

- Programming tutorial assignment on Cassandra
- Install Virtualbox (https://www.virtualbox.org) if you haven’t done so already
- Download the Cassandra OVA image for Virtualbox from http://goo.gl/D1iow8
- Go through the installation tutorial:
  - http://www.planetcassandra.org/install-cassandra-ova-on-virtualbox/
  - Go through the Developer Walkthrough: http://www.planetcassandra.org/create-a-keyspace-and-table
- Write a one paragraph summary of your experience, submit your 1 paragraph summary. We will grade these summaries.